Beyond the Pareto Efficient Frontier: Constraint Active Search for Multiobjective Experimental Design

Gustavo Malkomes, Bolong Cheng, Eric Lee, Michael McCourt

SigOpt, an Intel Company, San Francisco, CA, USA.

gustavo.malkomes@intel.com

ICML 2021

Beyond the Pareto Efficient Frontier: Constraint Active Search for Multiobjective Experimental Design




Gustavo Malkomes

Harvey Cheng

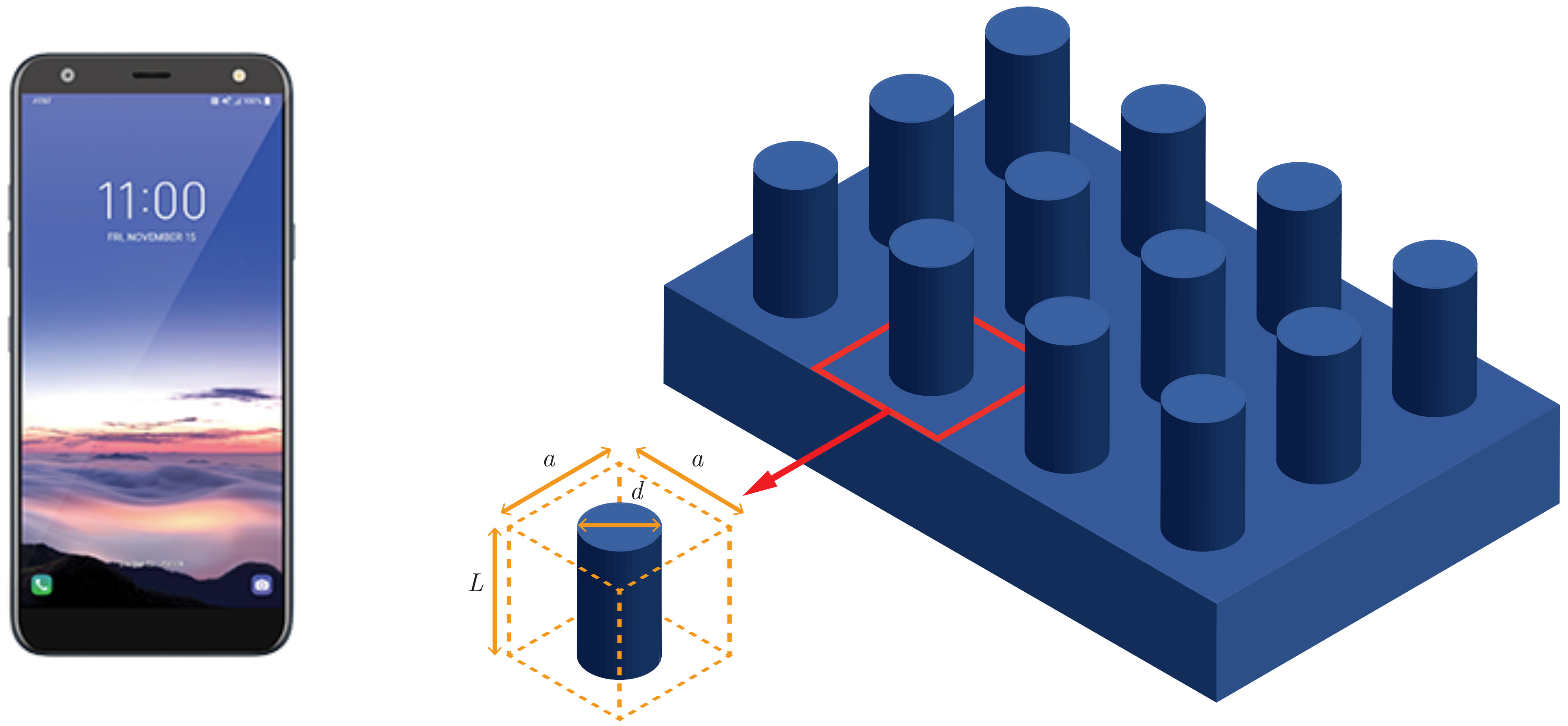
Eric Lee

Michael McCourt

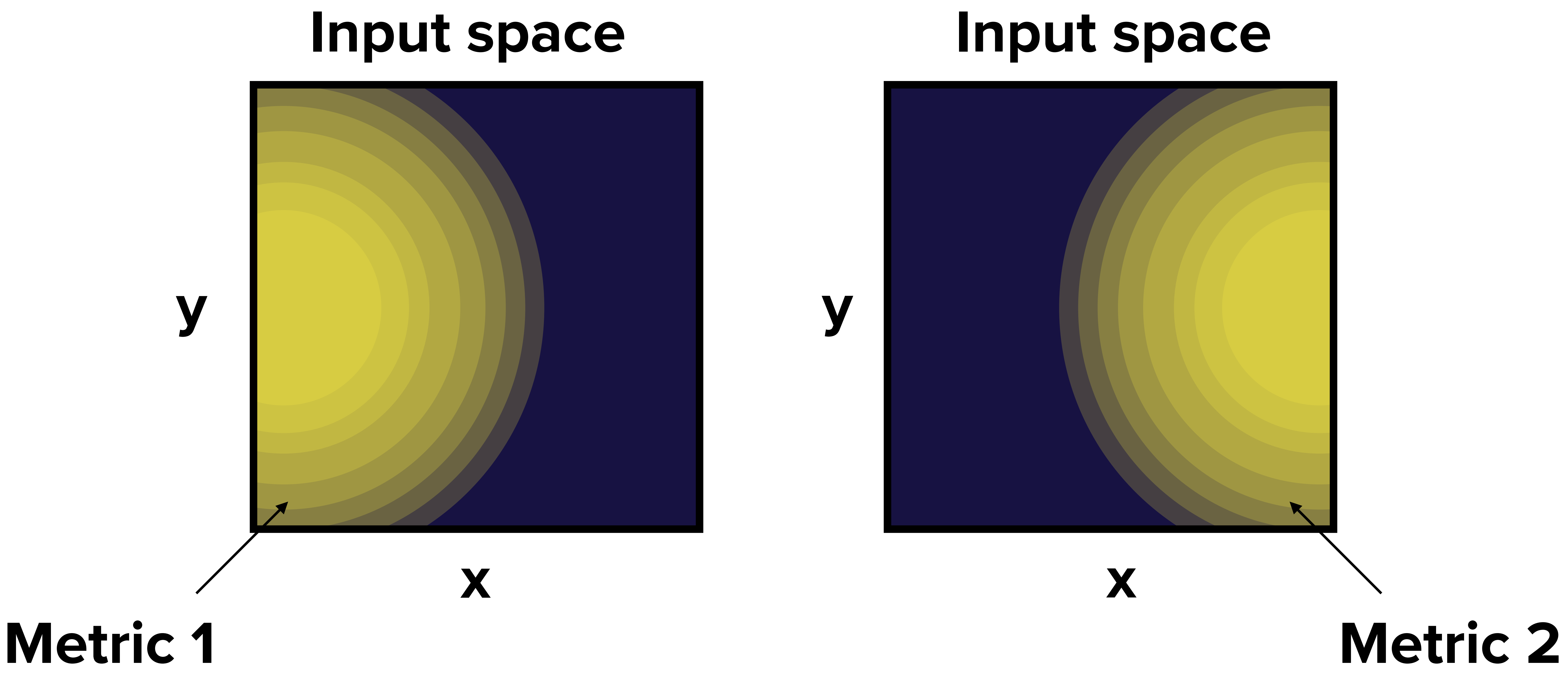
Accelerating Material Science Designs

Material scientists often faced challenging optimization problems

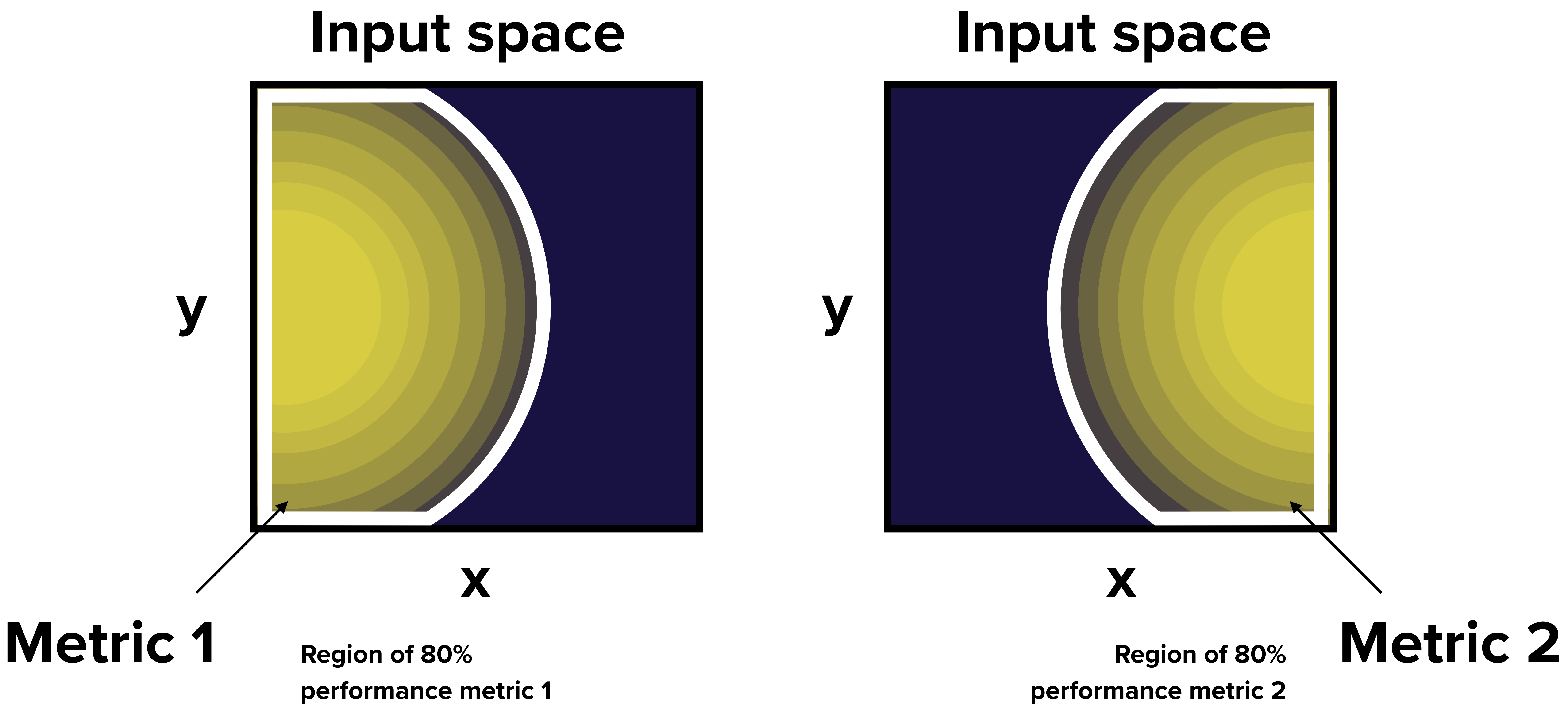
- **Multi-objective black-box problem** (minimize different reflection angles)
- **Practical metric constraints** exist on acceptable values for each objective
- **Limited and expensive budget** (each fabrication could take 5 days to execute)
- **Physical precision limitations** caused by the tools used in the experimentation



Balancing competing objectives

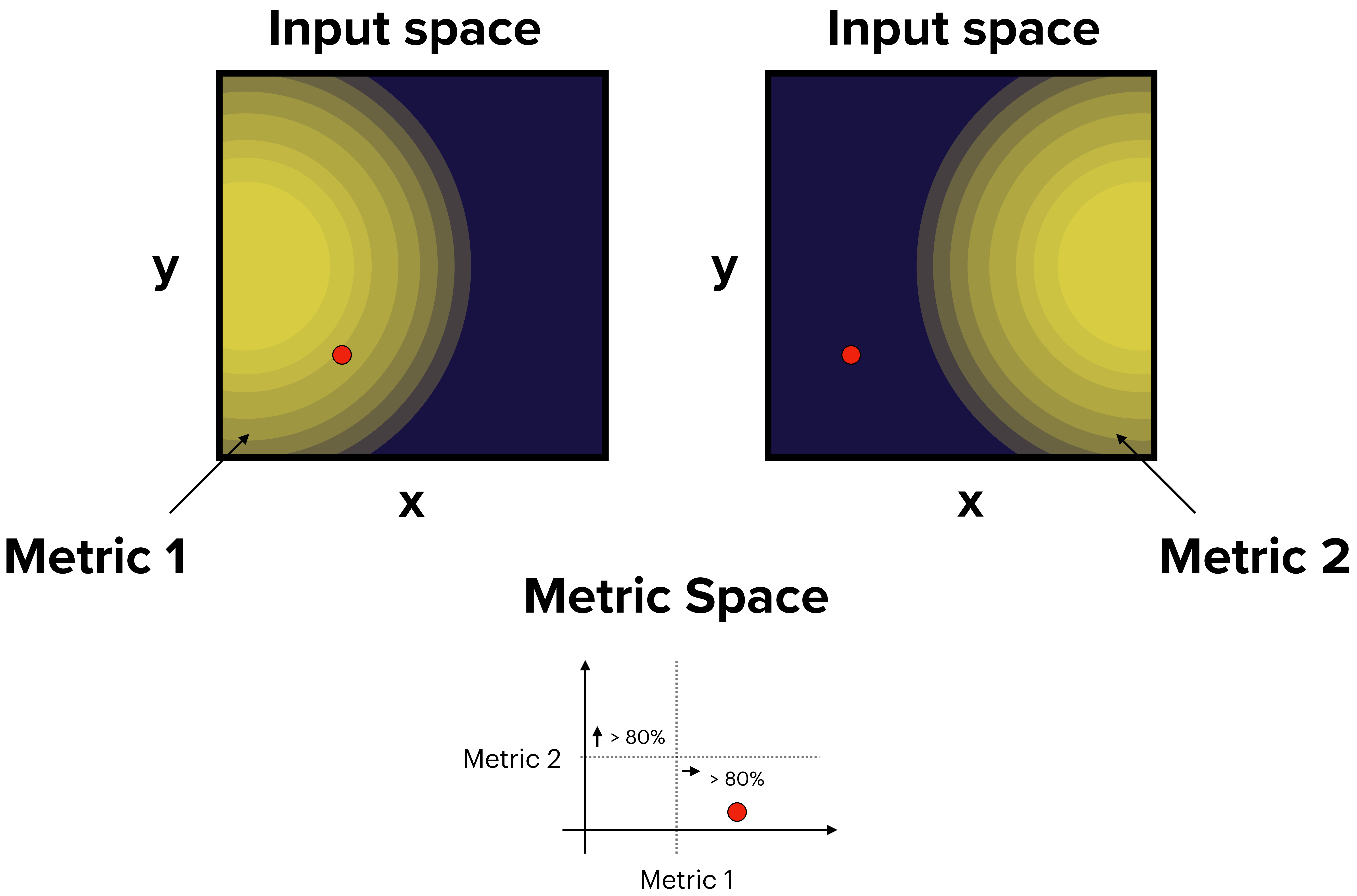


Practical metric constraints

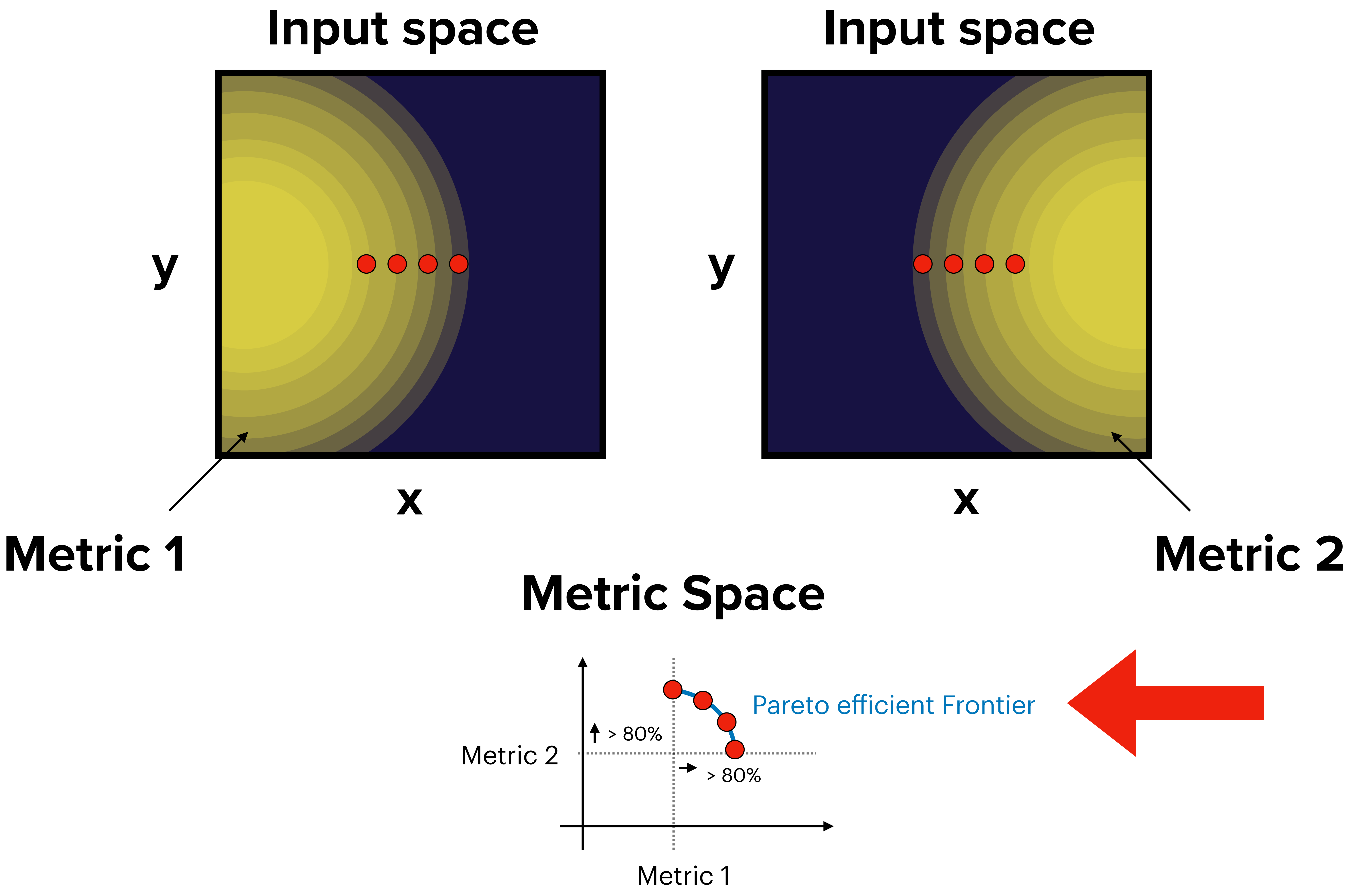


$$f_i(\mathbf{x}) \geq \tau_i \qquad f_i : \mathcal{X} \mapsto \mathbb{R}$$

Limited budget

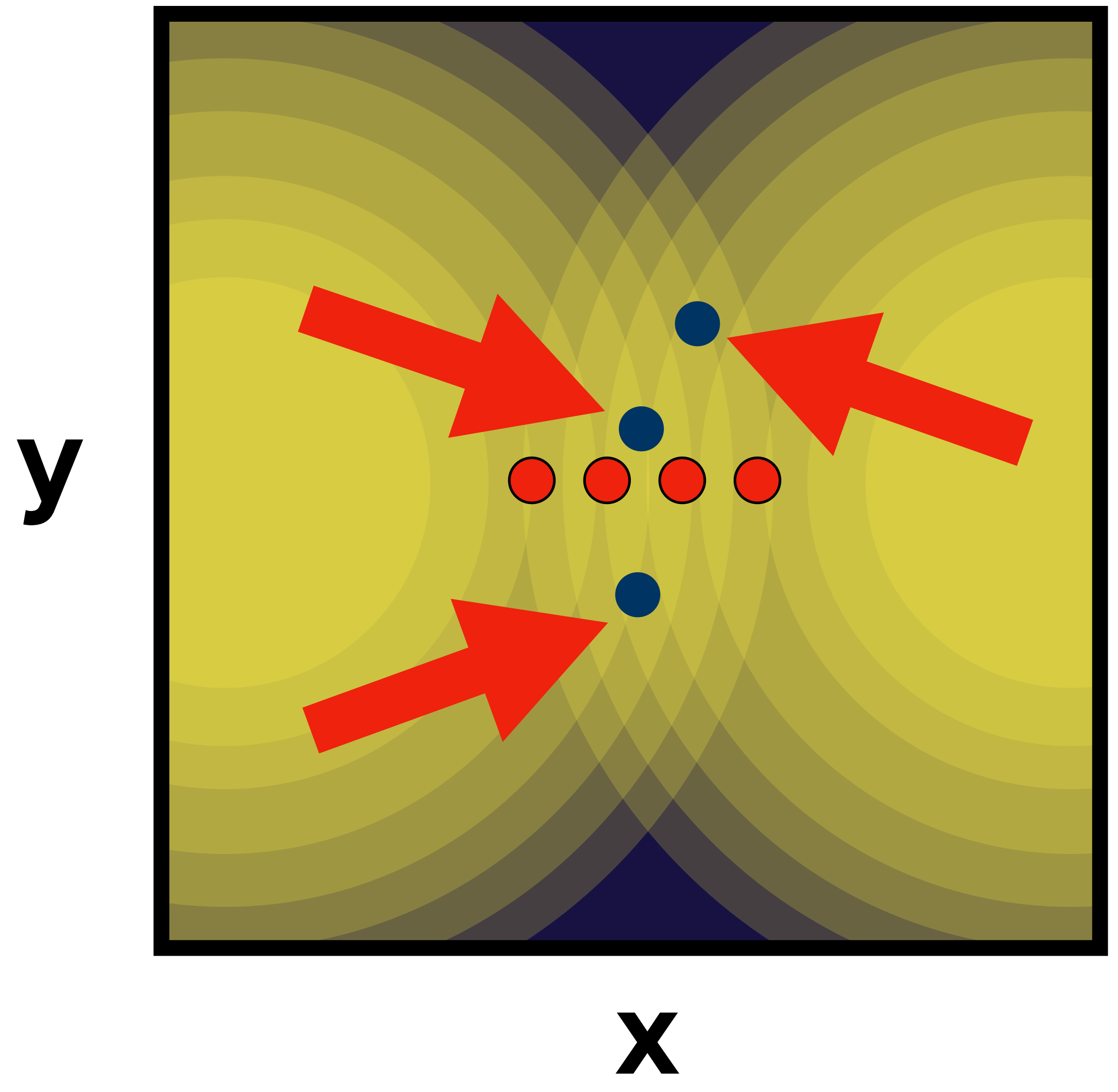


Intelligent multimetric optimization with constraints

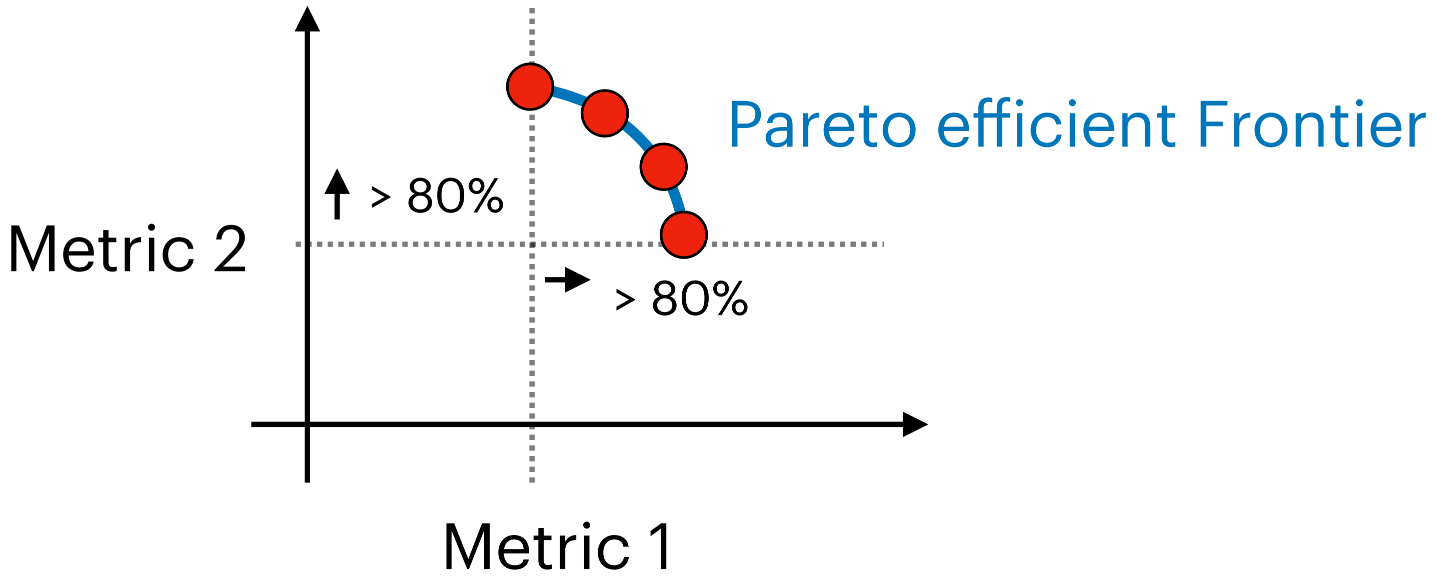


Intelligent multimetric optimization with constraints

Input space



Metric Space

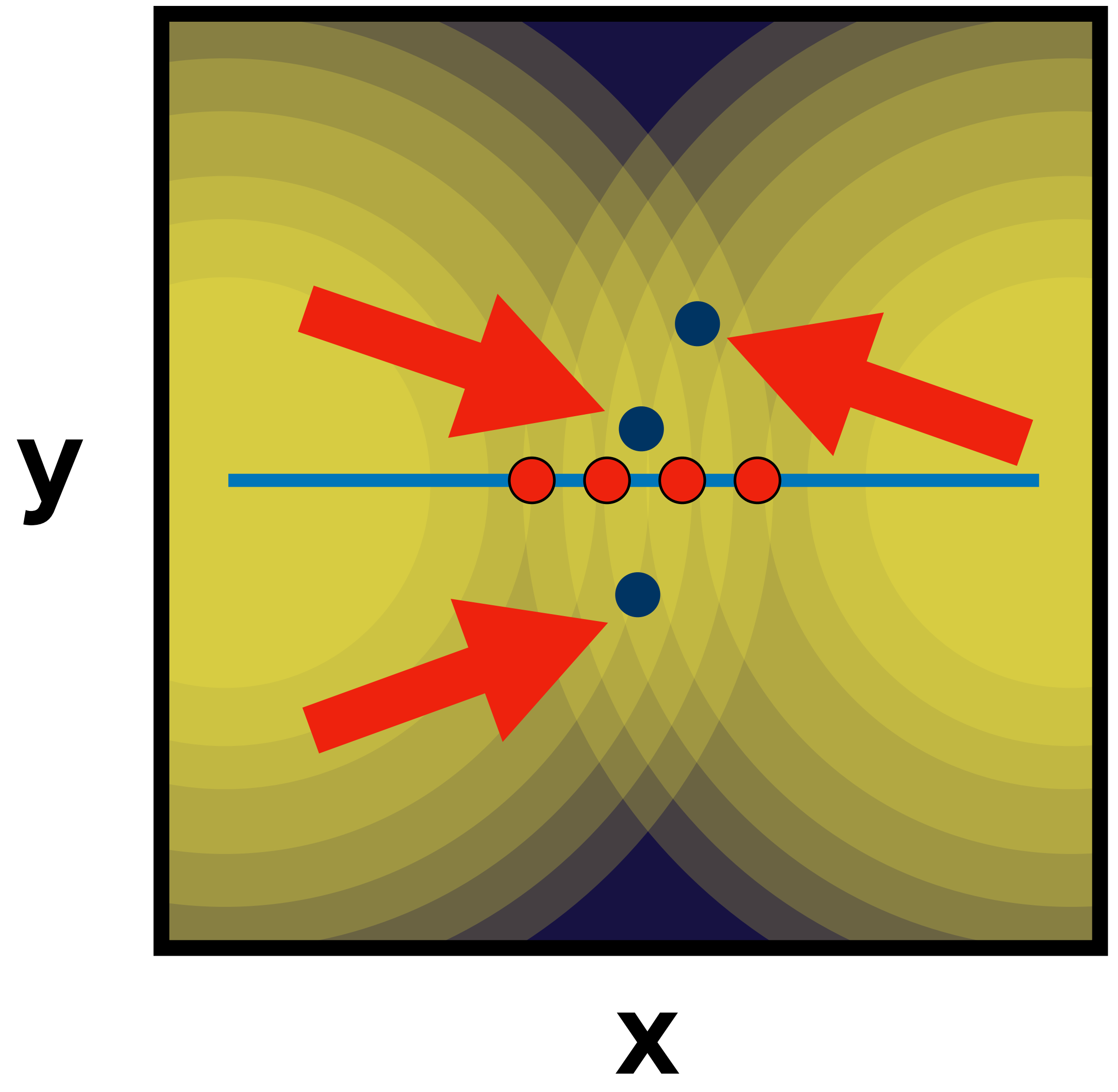


Dealing with precision limitations

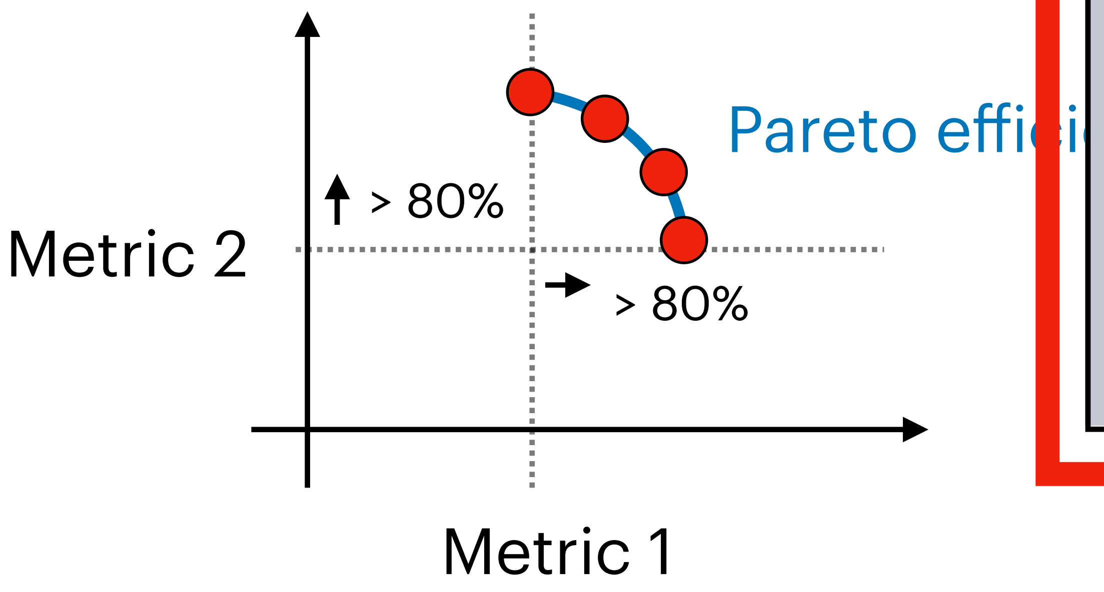
That's the optimal answer!

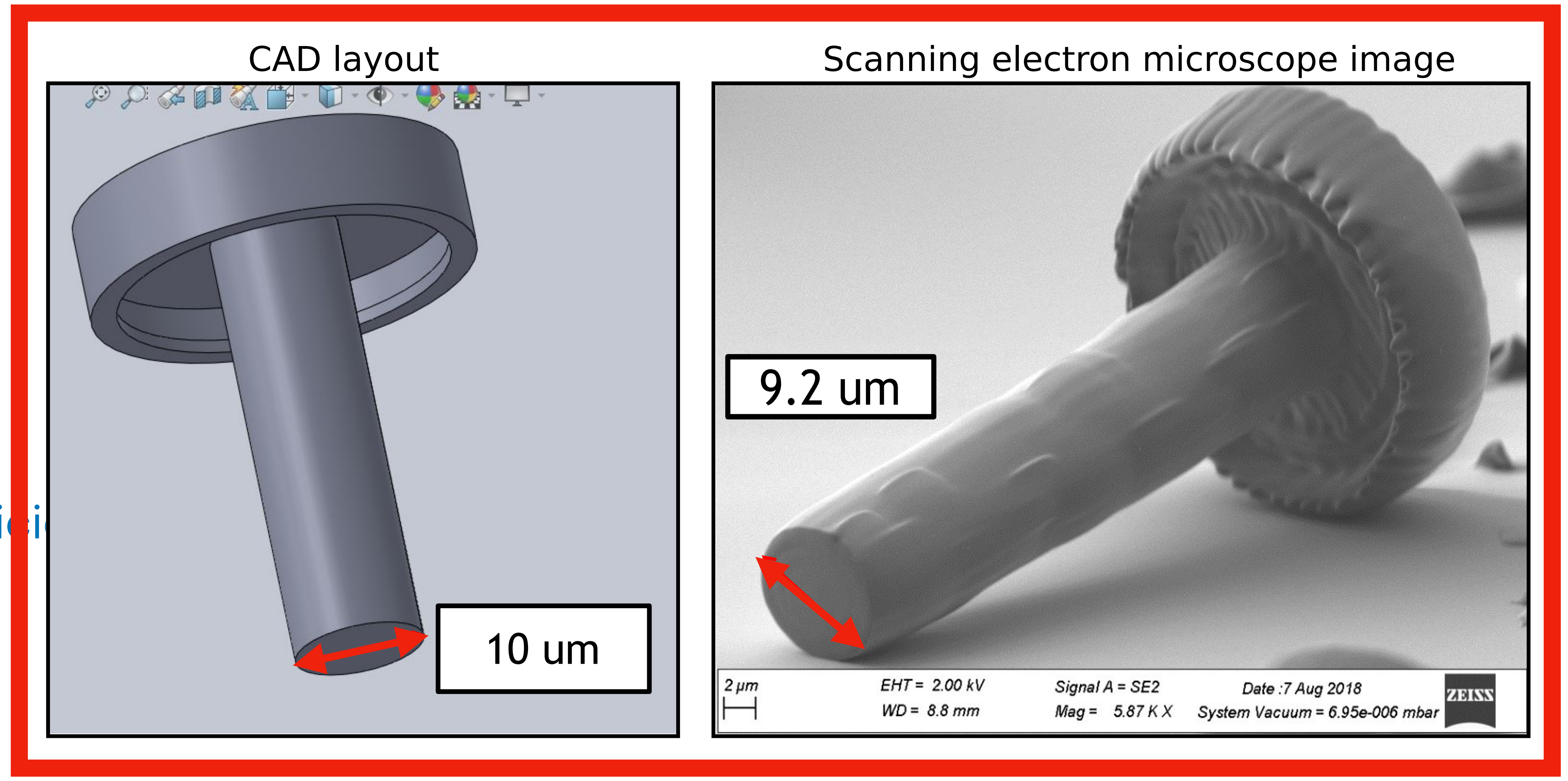
... during development
(numerical simulation)

Input space

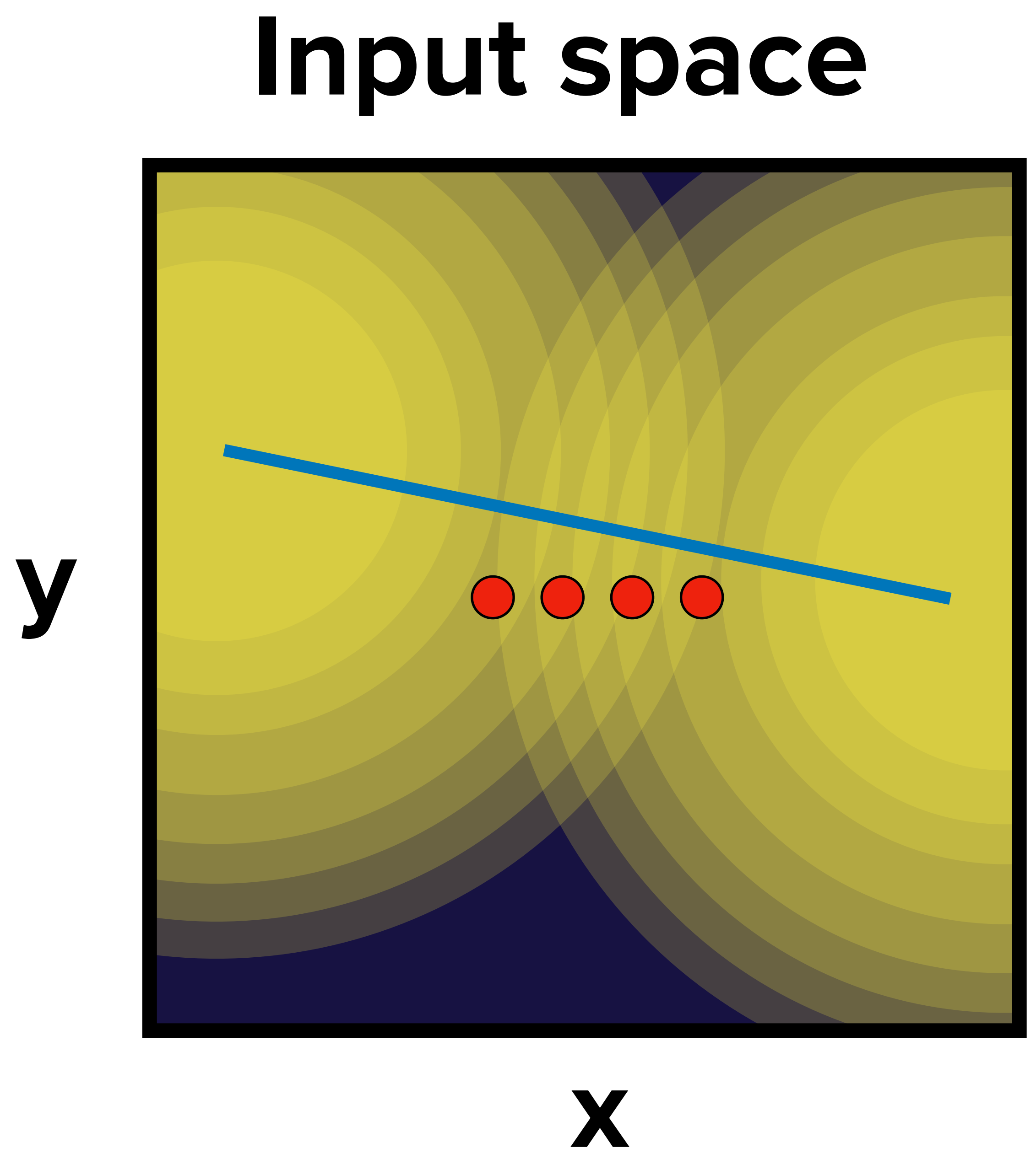


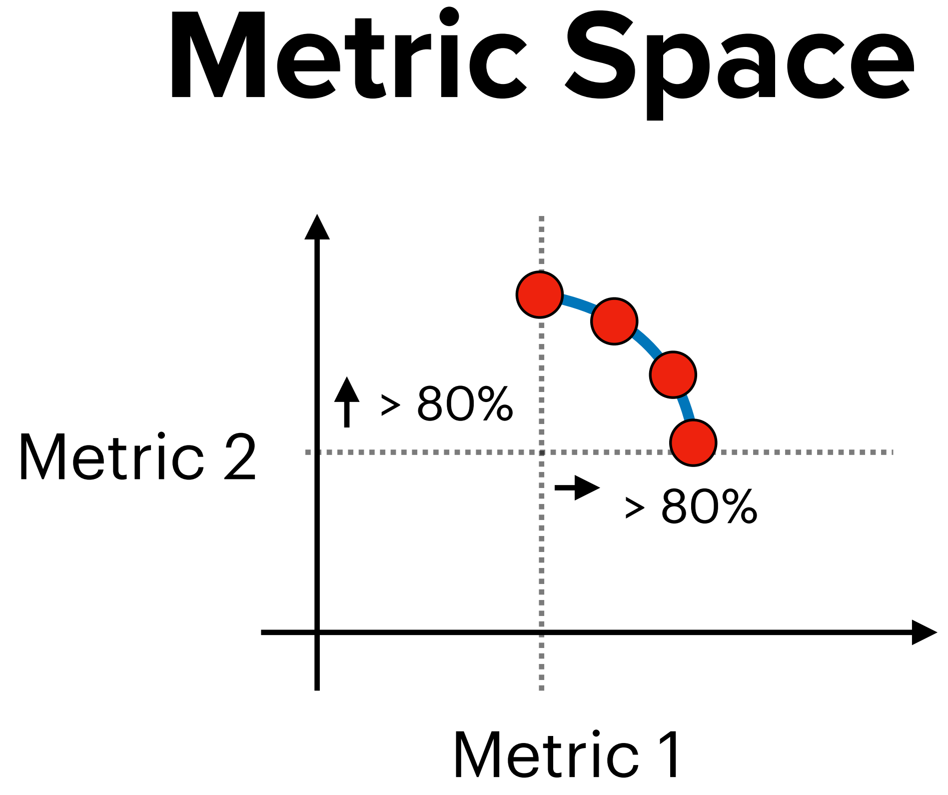
Metric Space





Discrepancy between development and production



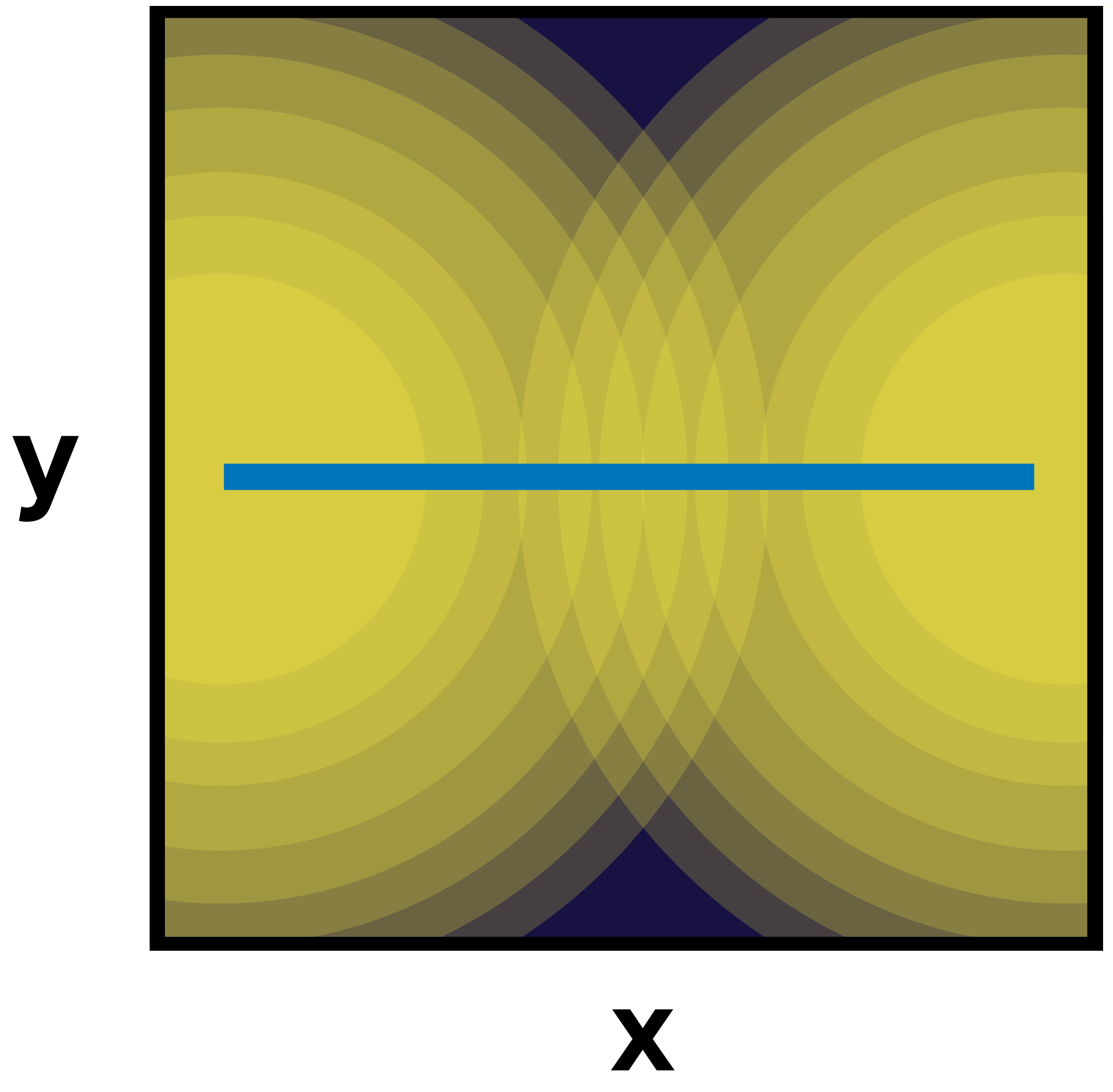


The “real” metrics are unknown a priori

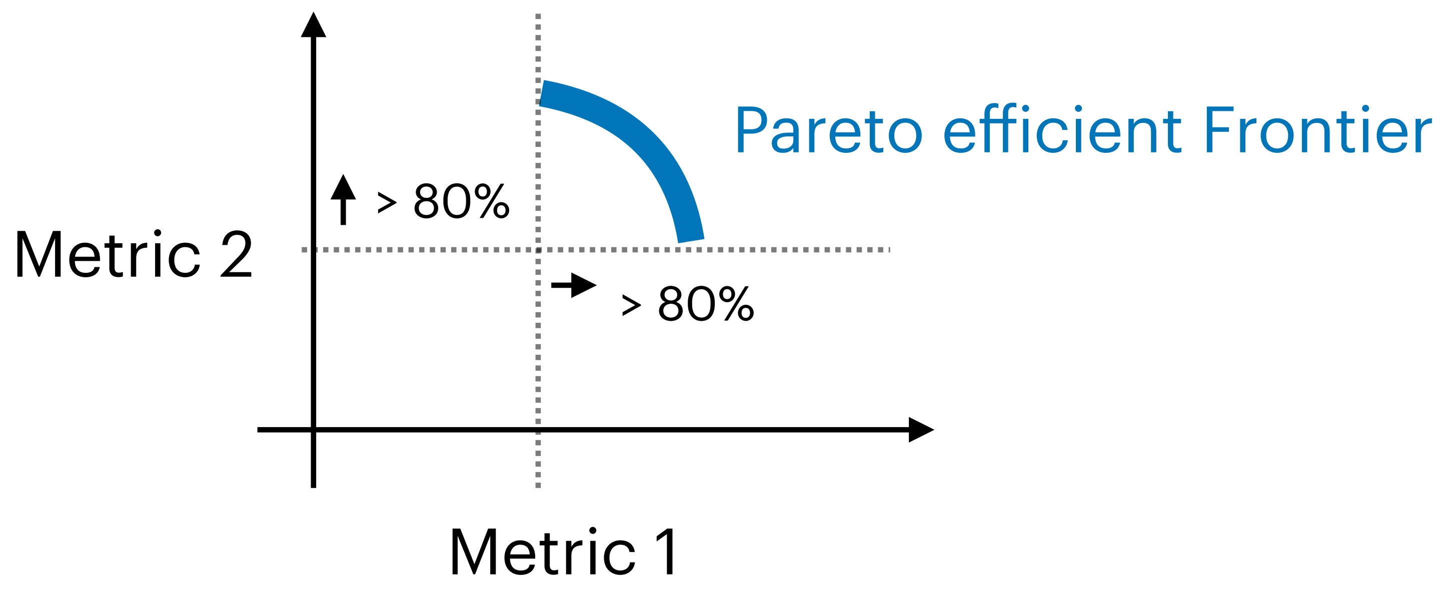
- Precision limitations
- Covariate shift
- Model mismatching

Multimetric formulations

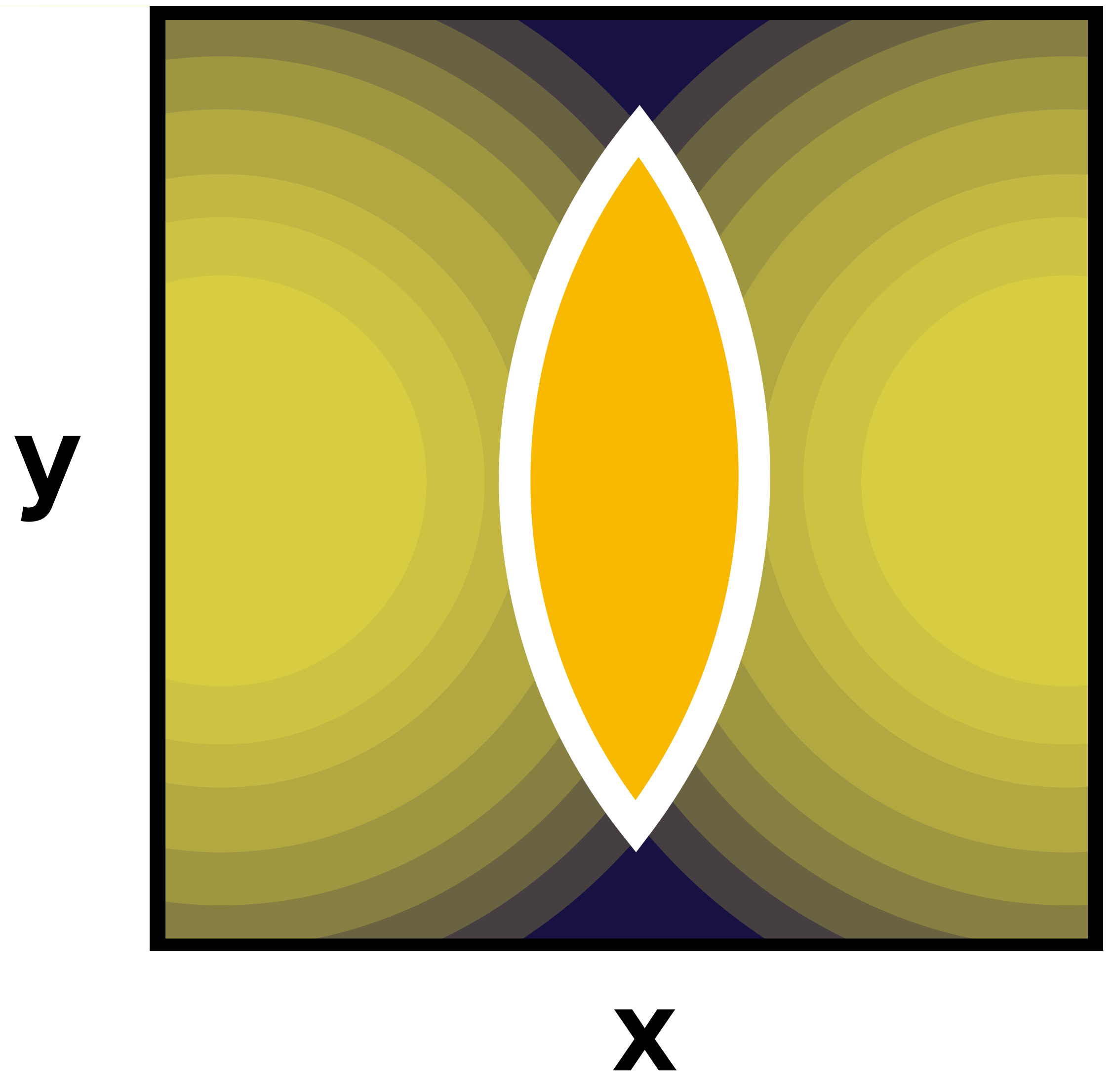
Multimetric optimization



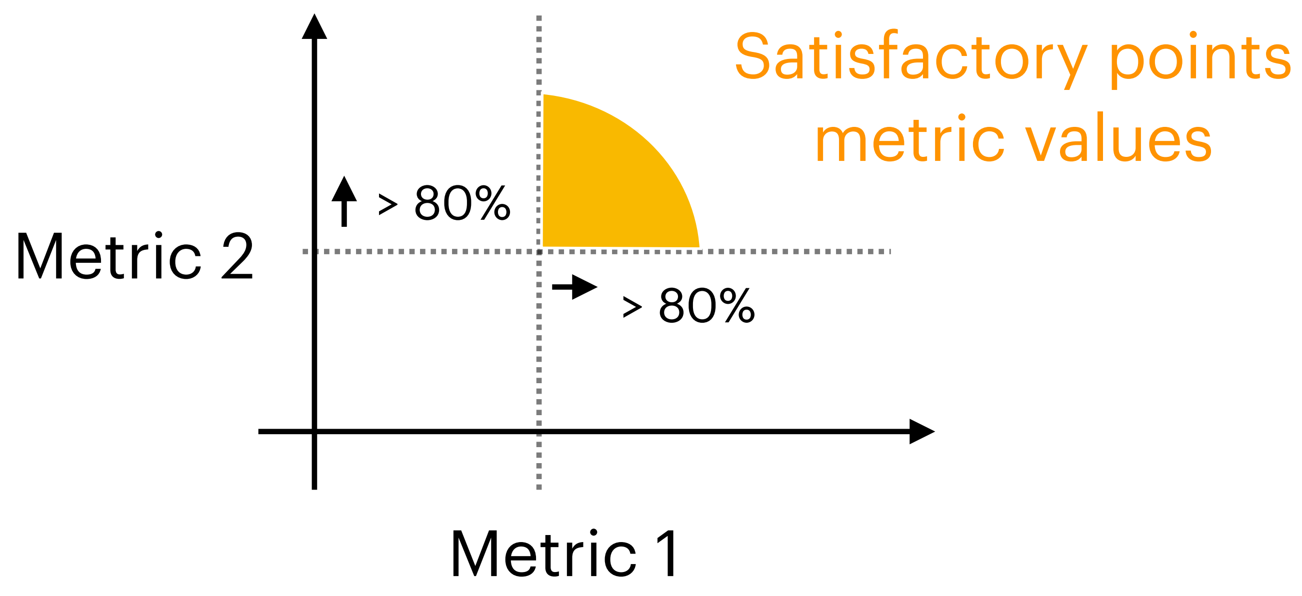
Metric Space



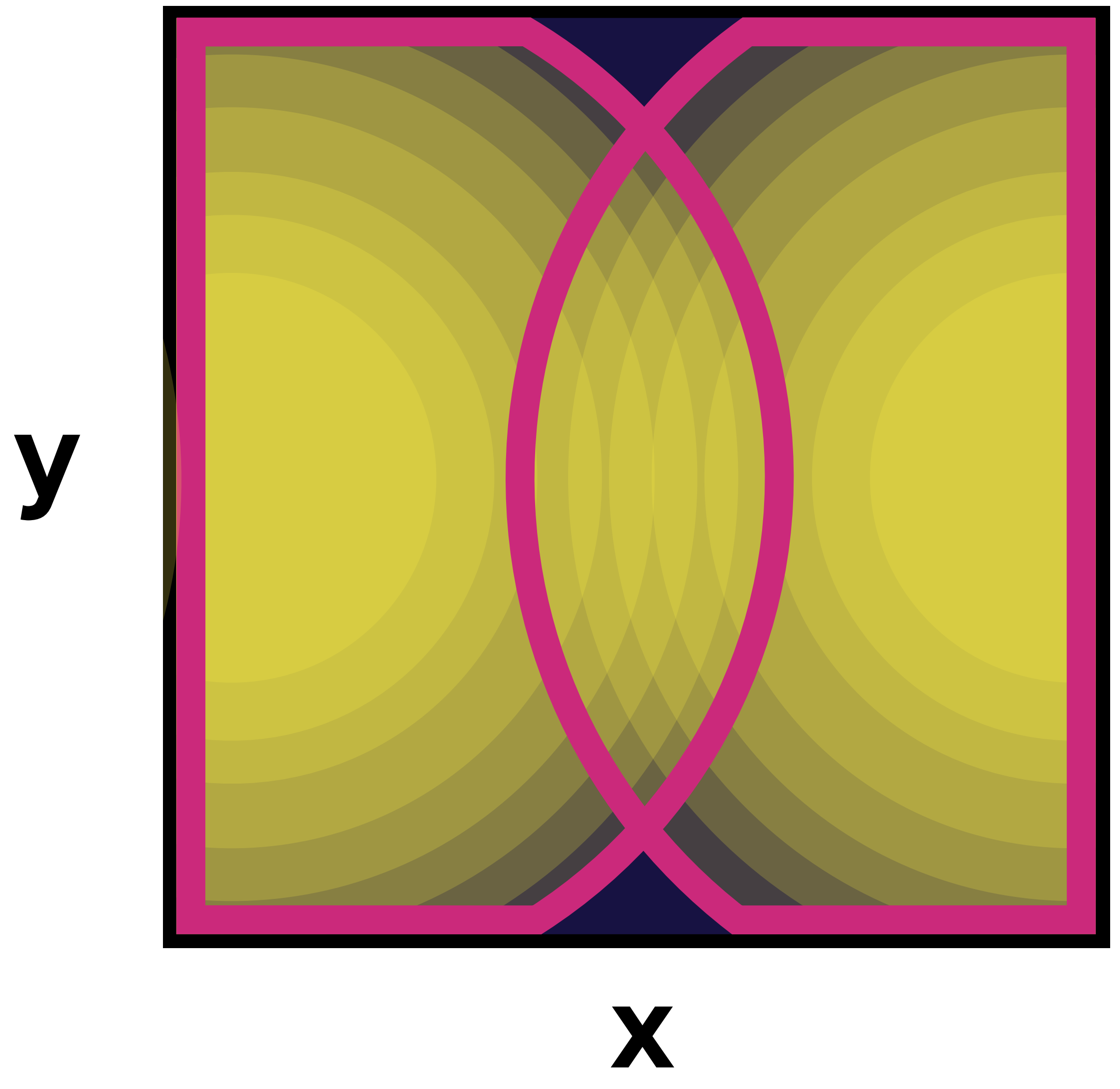
Constraint Active Search



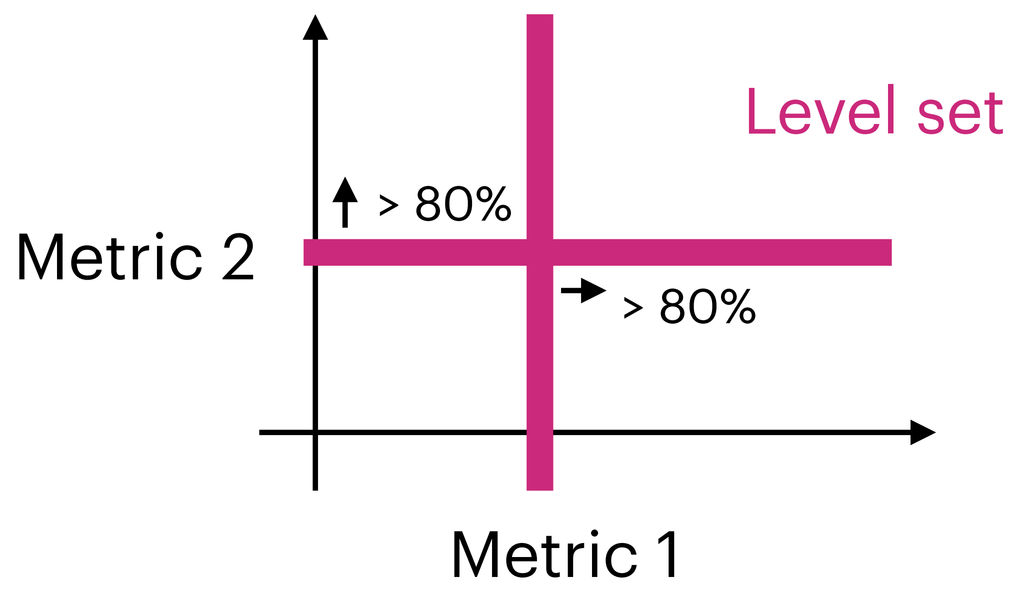
Metric Space



Level set estimation



Metric Space



Constraint Active Search

An alternative to the Pareto Frontier

Instead of performing multiobjective optimization, we solve a **search problem:**

Points in a input space

$$\mathbf{x} \in \mathcal{X}$$

Multiobjective functions $\mathcal{X} \mapsto \mathbb{R}$

$$f_1, f_2, \cdots, f_m$$

We propose soliciting desired minimum performance constraints to define a satisfactory region:

$$\boldsymbol{\tau} = [\tau_1, \tau_2, ..., \tau_m]^\top$$

Our goal is to search for a diverse set of configurations:

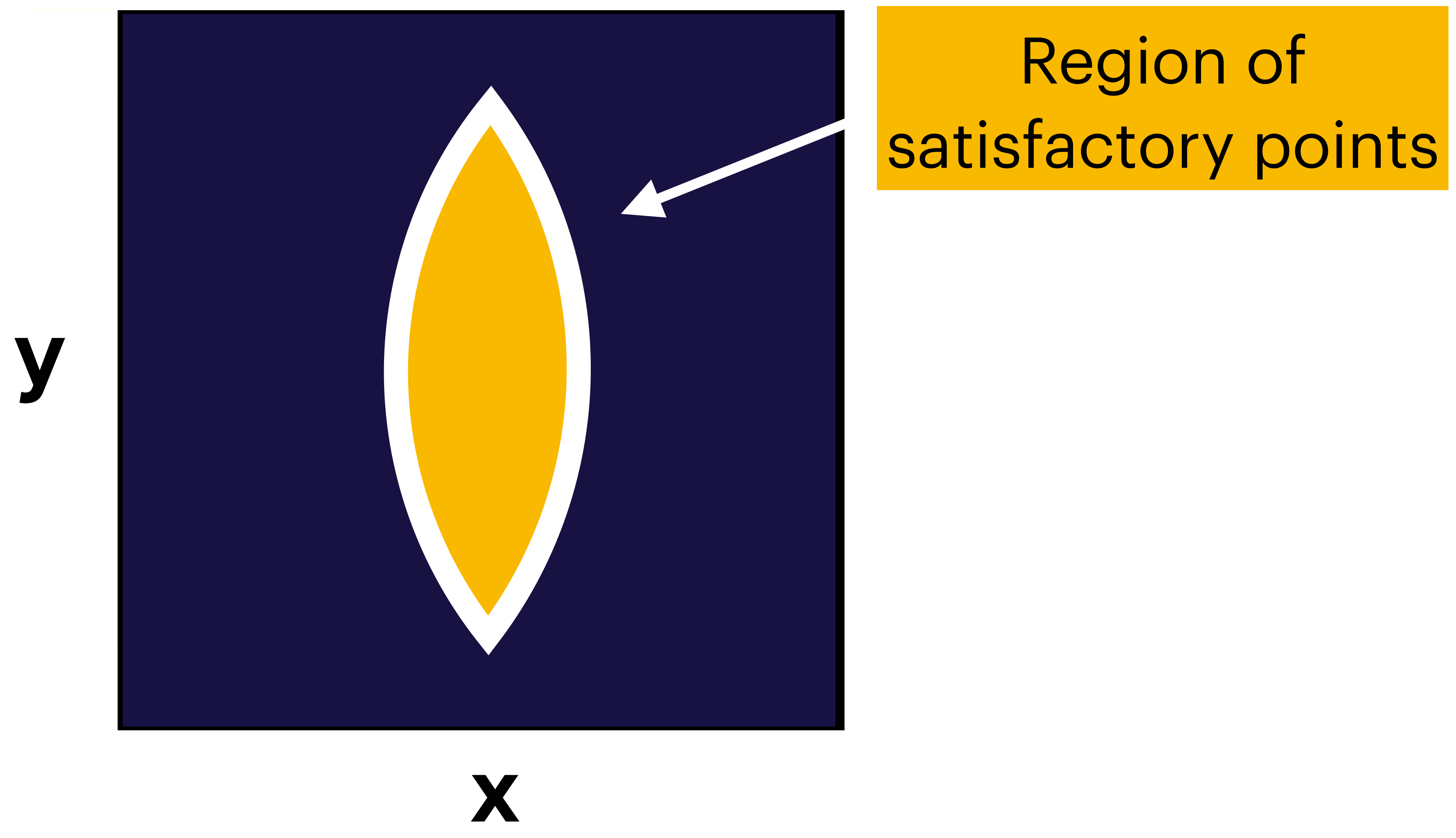
$$\mathcal{S} = \{\mathbf{x} \mid f(\mathbf{x}) \succeq \boldsymbol{\tau}\}$$

$$f(\mathbf{x}) \succeq \boldsymbol{\tau} := f_i(\mathbf{x}) \geq \tau_i$$

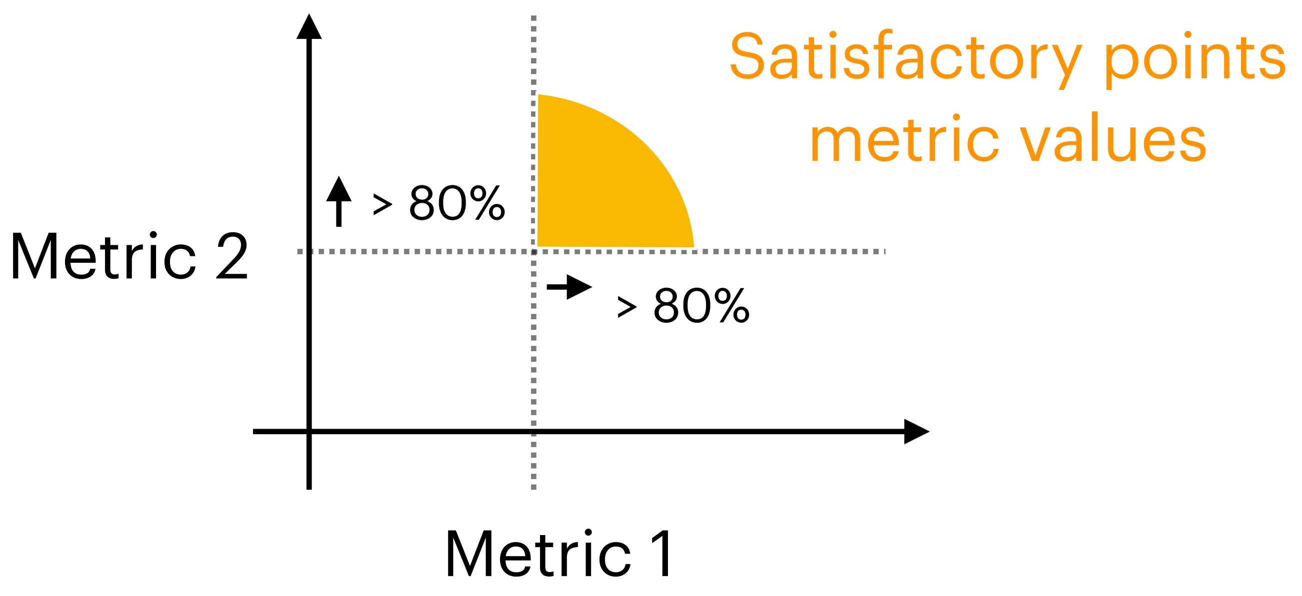
Low constraints: learns about the function everywhere **(experimental design)**

Set constraint to maximum value: equivalent to **Bayesian optimization**

Constraint Active Search



Metric Space



Constraint Active Search

Expected Coverage Improvement

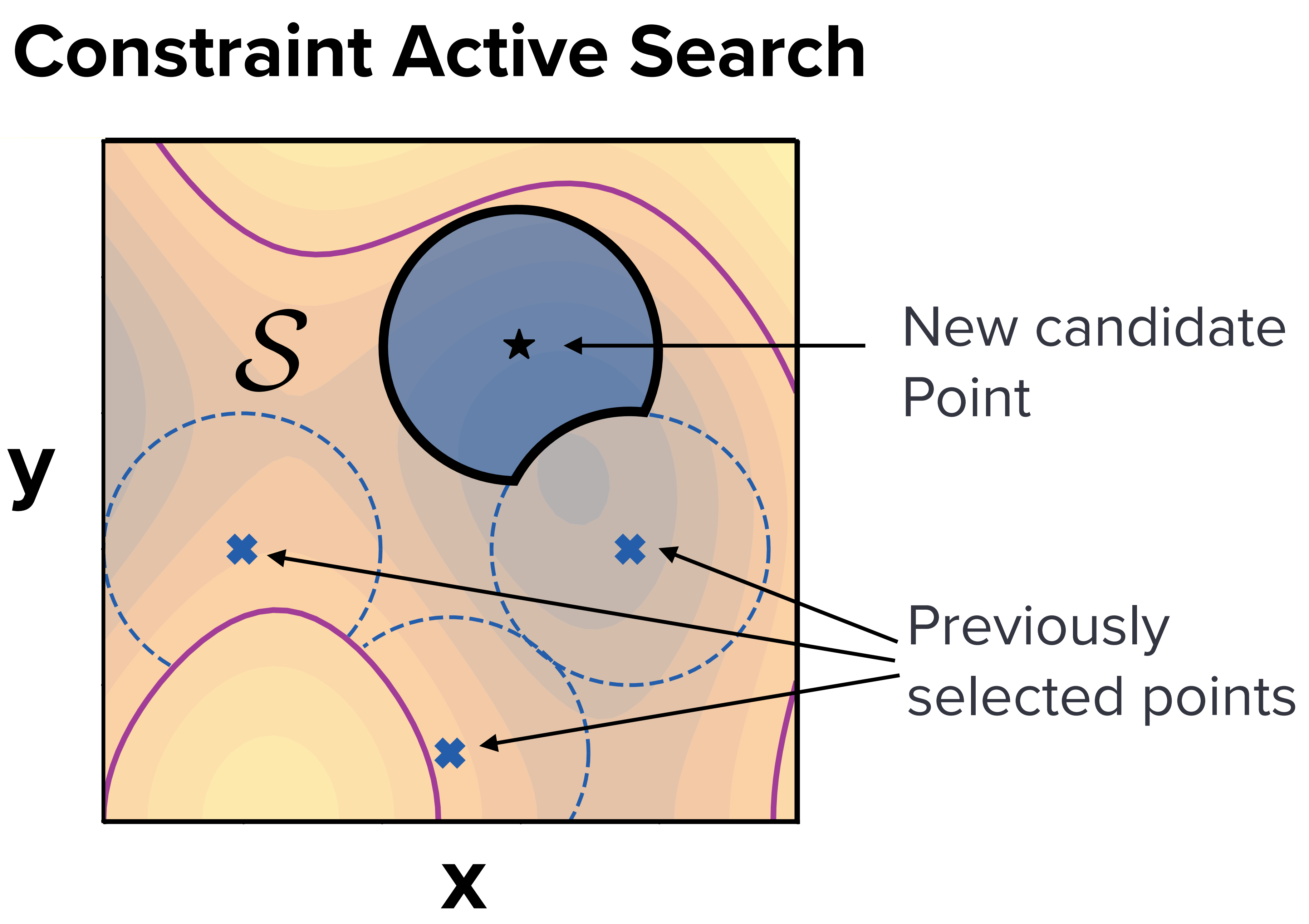
Select points that cover the satisfactory region given a resolution parameter r

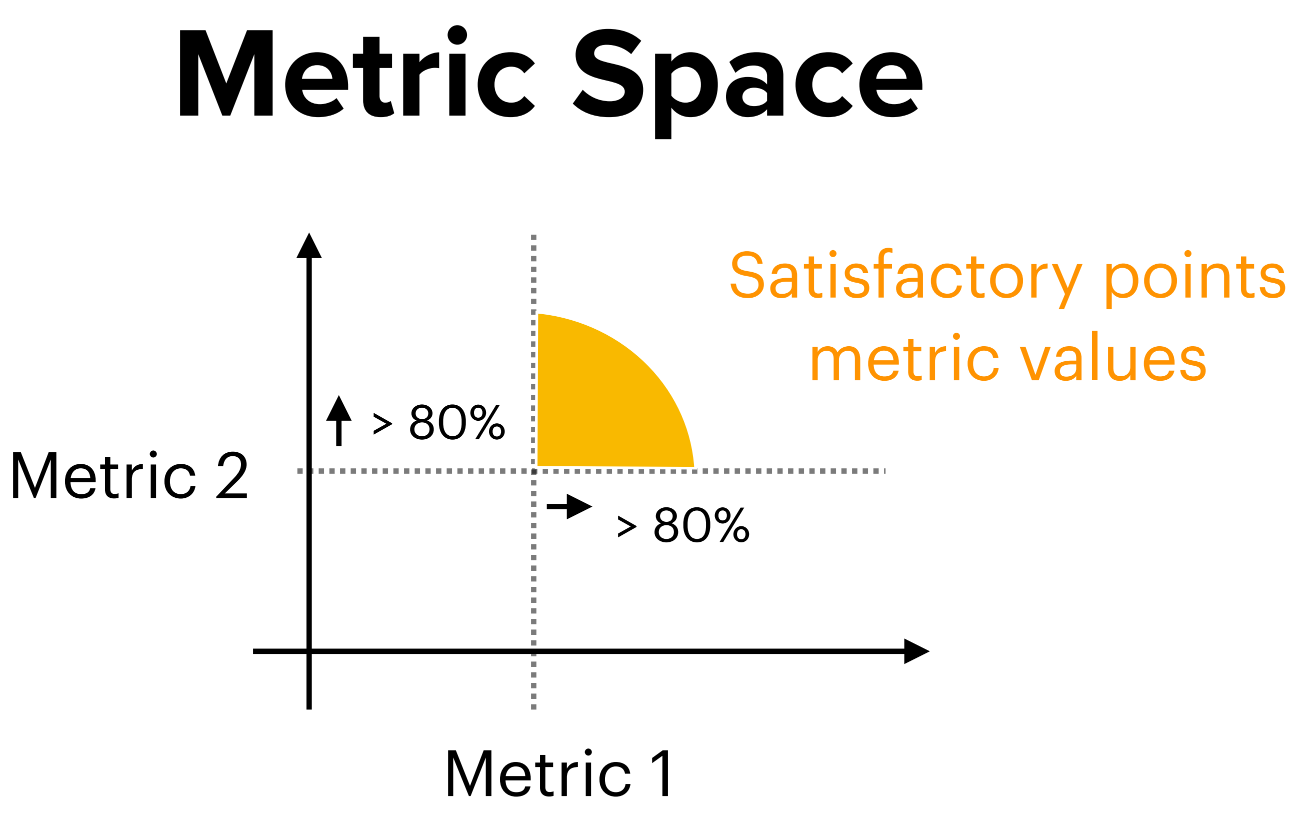
Our goal is to select points to cover the satisfactory region $\mathcal{S}$

We use Bayesian decision theory to derive an one-step optimal policy that covers the satisfactory region

ECI estimates the additional coverage increase given by the new point, considering the uncertainty around $\mathcal{S}$

Our policy selects the point with the highest expected increase in coverage, breaking ties by selecting the furthest observation from the selected points





Experiments

Evaluate multiobjective problems using multiple criteria

Fill distance

Radius of the largest empty ball we could place on the satisfactory region

$$\text{FILL}(\mathbf{X}, \mathcal{S}) = \sup_{\mathbf{x} \in \mathcal{S}} \min_{\mathbf{x}_j \in \mathbf{X}} d(\mathbf{x}_j, \mathbf{x})$$

Hypervolume

Volume of the region bounded by the Pareto points and the thresholds

Number of positives

Number of satisfactory points selected

Coverage recall

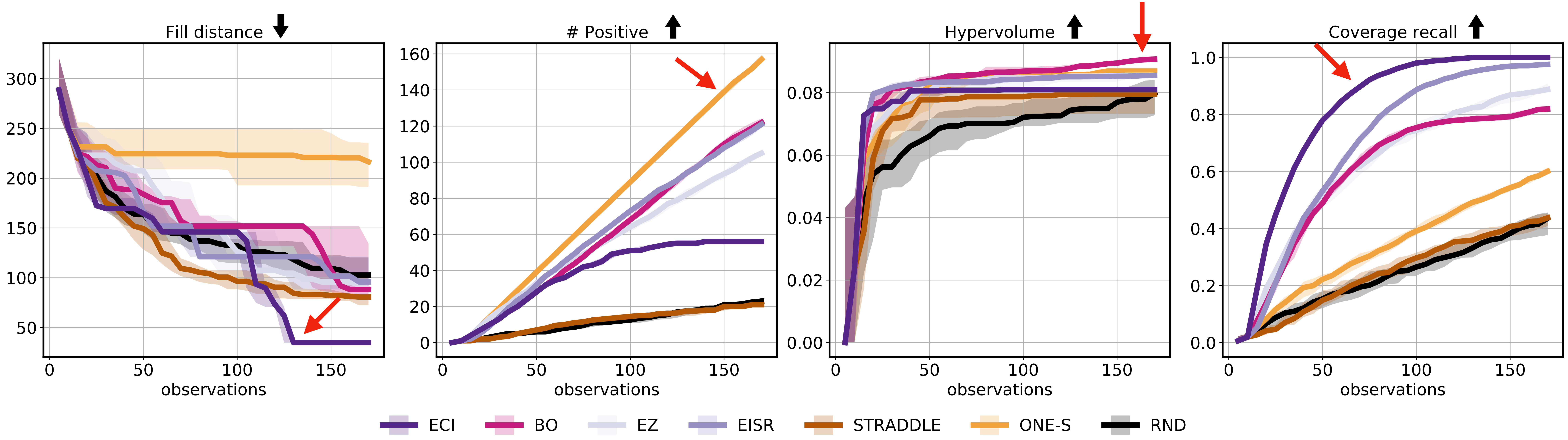
Induced volume of selected points inside the satisfactory region

Total of 11 experiments, 20 trials each

Several design and simulation domains: mechanical design, additive manufacturing, medical monitoring, and plasma physics

Experiments

Additive manufacturing example



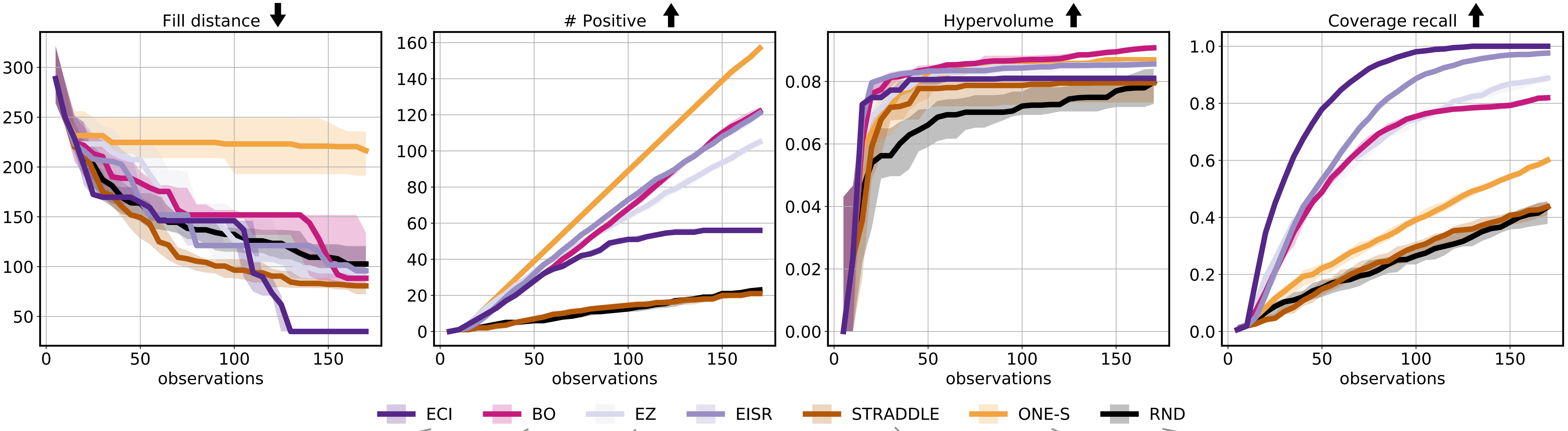
Our method ECI excels at finding **diverse** configurations inside the satisfactory region

Bayesian optimization consistently improves the **hypervolume** but is not as effective at diverse sampling

Similarly, **active search**, which greedily maximizes the **number of positive points**, performs best at this metric but the samples are not diverse

Experiments

Additive manufacturing example



Expected Coverage Increase **(our method)**

Bayesian Optimization, ε-constraint (Haghanifar et al., 2020)

Mutual information between output variable and probability of a point being above the threshold (level set estimation baseline)

Entropy inside the satisfactory region, rewards points with high uncertainty if they are above the threshold **(our proposed baseline)**

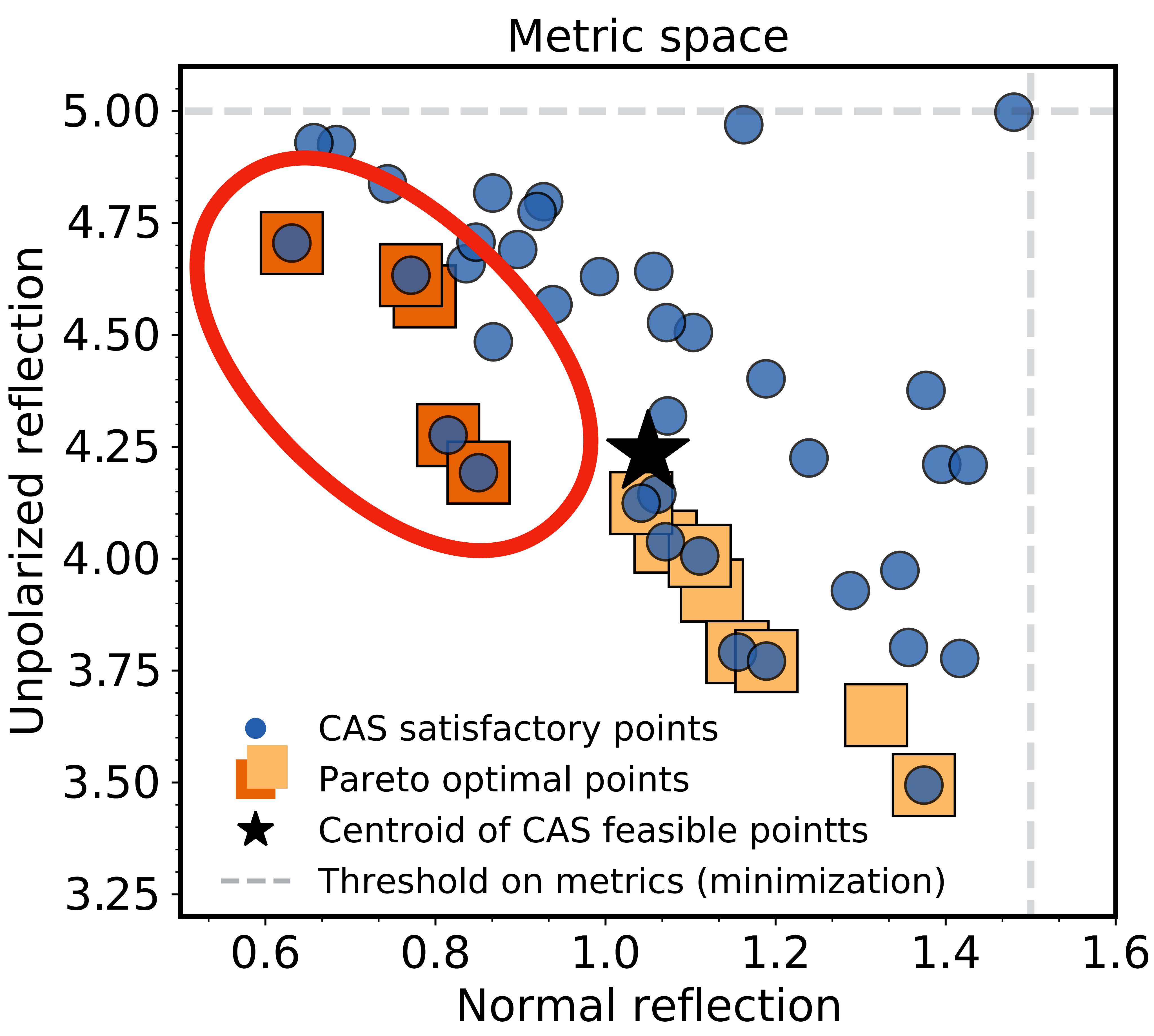
Straddle heuristic adapted from Bryan et Al., (2005) (level set estimation baseline)

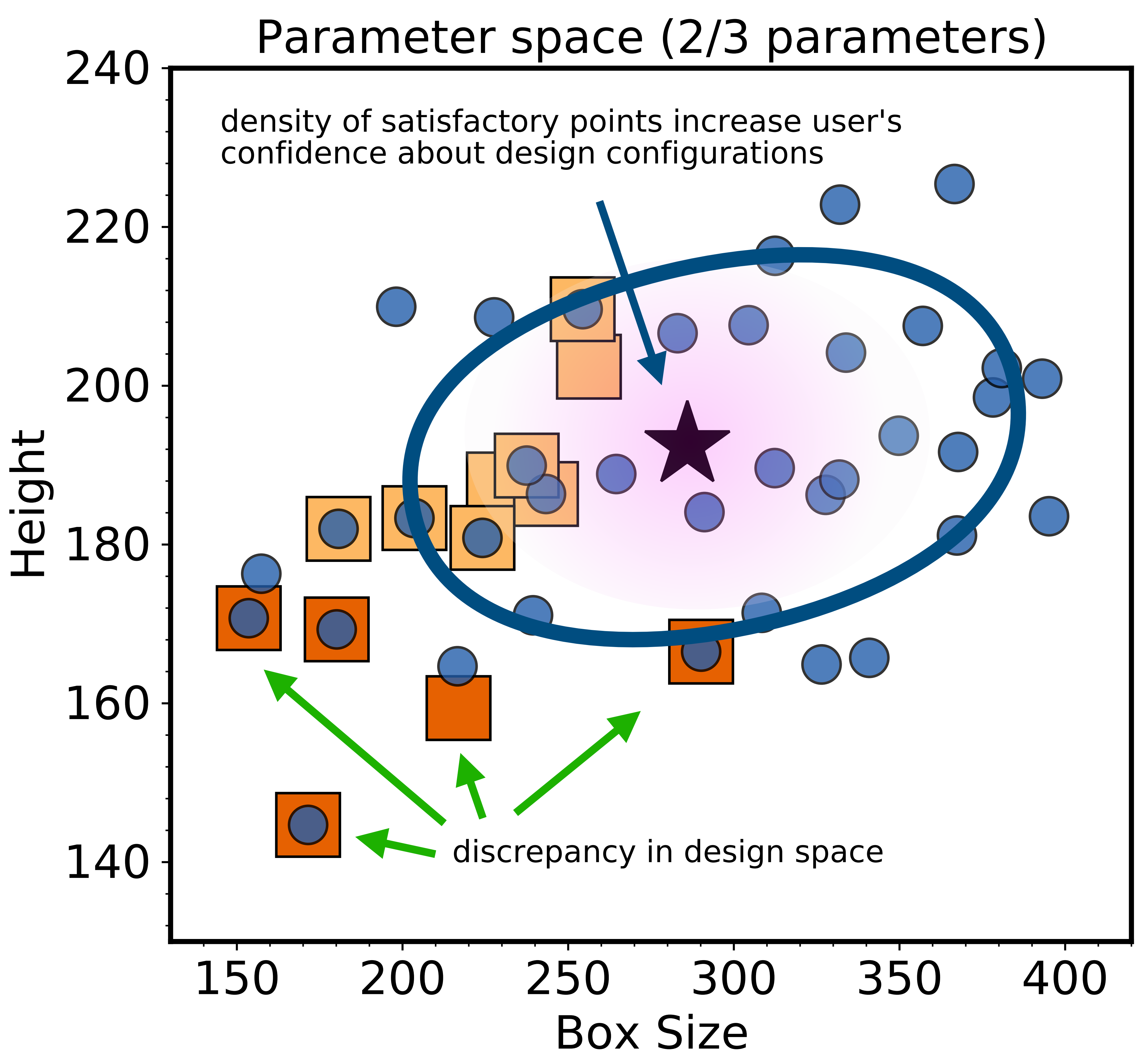
One step active search (Garnett et al., 2012)

Random

Please check our paper for more details

Beyond the Pareto efficient Frontier





Key contributions

- We introduce a paradigm for **multi-objective black-box problems**, which we call **constraint active search (CAS)**. CAS can be seen as a generalization of Experimental Design and Bayesian Optimization

- We develop an algorithm called **expected coverage improvement (ECI)**. ECI focus on searching **diverse samples** that satisfy the **objective constraints**. We also provide theoretical analysis on the sample diversity of ECI

- Theoretical properties of this strategy include a **constant approximation ratio** to the optimal **sample diversity (fill distance)**

- We compare ECI to various baselines on a suite of synthetic multiobjective design benchmarks as well as **real-world multiobjective design and simulation applications** in materials science, medical monitoring, and plasma physics

Thank you!

contact: gustavo.malkomes@intel.com

Σ intel